AFRL-AFOSR-VA-TR-2016-0173

Sample-starved large scale network analysis

Alfred Hero Iii
UNIVERSITY OF MICHIGAN

05/05/2016
Final Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ RTA2
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
28/04/2016	Final	02/01/2013 - 01/31/2016

4. TITLE AND SUBTITLE
Sample-Starved Large Scale Network Analysis

5a. CONTRACT NUMBER
FA9550-13-1-0043

5b. GRANT NUMBER
AFOSR/PKR2

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Alfred O. Hero, University of Michigan
Bala Rajaratnam, Stanford University

5d. PROJECT NUMBER
F032773

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Regents of the University of Michigan
Division of Research Development and Administration
503 Thompson St.
Ann Arbor, Mi 48109-1340

8. PERFORMING ORGANIZATION REPORT NUMBER
12-PAF03300

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
USAF, AFRL DUNS143574726
AF Office of Scientific Research
875 N. Randolph St. Room 3112
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
USAF, AFLR, AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
FA9550

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution Unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
In this research project we developed correlation mining methods to answer the following fundamental question about complex networks: What are the fundamental limits on the amount of information that can be inferred about a network from a small number n of indirect empirical observations? In these terms, the overall objective was to develop algorithms and establish performance limits for mining information from correlation networks. The focus was on the sample starved regime arises when the number of variables (columns of the correlation matrix) is of the same order or larger than the number of observations available to estimate or detect patterns in the matrix.

15. SUBJECT TERMS
multivariate dependency structure
multivariate spatio-temporal prediction
anomaly detection

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			A. O. Hero
					19b. TELEPHONE NUMBER *(Include area code)*
Unclassified	Unclassified	Unclassified	UU	1	734-763-0564

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE. Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE. State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATE COVERED. Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

4. TITLE. Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER. Enter all contract numbers as they appear in the report, e.g. F33315-86-C-5169.

5b. GRANT NUMBER. Enter all grant numbers as they appear in the report. e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER. Enter all program element numbers as they appear in the report, e.g. 61101A.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER. Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER. Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/ monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.

AFOSR Grant: Sample starved inference in large scale networks

AFOSR Grant Number: FA9550-13-1-0043
Grant period: 2/1/13 to 1/31/16

Final Report

Personnel:

Alfred O Hero, III (PI)
University of Michigan
Ann Arbor, MI 48109-2122
hero@umich.edu

Bala Rajaratnam (co-PI)
Stanford University
Stanford CA
brajarat@stanford.edu

Abstract

In this research project we developed correlation mining methods to answer the following fundamental question about complex networks:

> What are the fundamental limits on the amount of information that can be inferred about a network from a small number n of indirect empirical observations?

In these terms, the overall objective was to develop algorithms and establish performance limits for mining information from correlation networks. The focus was on the sample starved regime arises when the number of variables (columns of the correlation matrix) is of the same order or larger than the number of observations available to estimate or detect patterns in the matrix. A new framework was developed to answer the above question based on spherical Gram matrices for inferring dependency structure of large networks from limited and/or incomplete sample observations of network behavior. The geometrical and statistical properties of these matrices was studied in the finite sample regime and in the asymptotic limit as numbers of samples and/or nodes become large. These properties led to quantification of fundamental performance tradeoffs and gave insights into phase transitions and convergence rates for inferring dependencies in network data. The theory was applied to practical complex network inference tasks including: online prediction, network variable selection and error controlled topology discovery.

1 Project Overview

This research project addressed the following important question: what are the fundamental properties of a network of interacting variables that can be accurately estimated from a small number of measurements? Properties of interest include collections of variables that are hubs, cliques and separators in the dependency graph associated with the network. This question lies as the foundation of network science yet had not been previously addressed in the context of "sample starved" regimes where large network size, limited node accessibility, or fast network dynamics make collection of large amounts of relevant data infeasible. On the other hand, almost all existing approaches to answering this question rely explicitly or implicitly on asymptotically large sample size assumptions. Indeed, network learning, inference and coding methods are commonly evaluated using strong law of large numbers, central limit theorem, and concentration inequalities. All of these methods give useful information about performance only as the number of samples (n) goes to infinity. However, in sample starved situations where n is small, these classical laws of large numbers simply do not apply and therefore asymptotic results are not useful. The researchl focused on sample-starved inference and structure discovery problems for networks of inter-dependent variables. We developed performance predictions that accommodated small n regimes and this led to a new small n theory. This theory applies to incomplete observations that inevitably result in large networks of variables. We also developed scalable and accurate algorithms for estimating the graphical model of multivariate dependency structure.

The question of sample starved structure discovery was formulated in the framework of multivariate dependency networks, also known as graphical models. One of the fundamental measures of multivariate dependencies is the covariance or correlation matrix. The graphical model associated with linear dependency is determined by the thresholded correlation matrix. Estimating the correlation matrix is a fundamental problem and it plays a crucial role in many inferential and data analysis methods. Principal component analysis (PCA), multivariate analysis of variance(MANOVA), classification via linear/quadratic discriminant analysis (LDA/QDA), canonical correlation analysis (CCA) and partial least squares (PLS) all require estimating the covariance matrix, its inverse (referred to as as the concentration or precision matrix) or some other function of the elements of the covariance matrix. These data analysis methods frequently arise in network inference tasks such as: network anomaly detection, network intrusion detection, social network community detection, social networks, and network tomography. In this context, the sample starved regime arises when the number of variables (columns of the correlation matrix) is significantly larger than the number of observations available to estimate the elements of the matrix. In this project we developed a general random matrix framework that applies to the broad class of estimators based on thresholded sample correlation and pseudo-inverse correlation matrices

2 Accomplishments/New Findings

Our accomplshments fall into several areas listed below, and are discussed in more detail in the sequel.

1. Foundational principles for large scale inference on structure of covariance
2. Predictive correlation screening with resource constraints

3. Local hub screening in a correlation network
4. Non-convex sample-starved estimation of sparse inverse covariance matrices
5. Positivity invariance of thresholded correlation matrices
6. Applications to materials science

2.1 Foundational principles for large scale inference on structure of covariance

We developed general principles for reliable inference of covariance structure in the Big Data setting? A book chapter (A.O. Hero and B. Rajaratnam, "Large scale correlation mining for biomolecular network discovery," in Big Data Over Networks, 2015) and a journal article (A.O. Hero and B. Rajaratnam, "Foundational principles for large scale inference: Illustrations through correlation mining," IEEE Proceedings.vol. 105, no. 1, pp. 93-110, Jan. 2016) presents these principles in a concise but accessible format. These principles are applicable to large-scale complex network applications arising genomics, connectomics, eco-informatics, and elsewhere, where the data set is often variable rich but sample starved: a regime where the number n of acquired samples (statistical replicates) is far fewer than the number p of observed variables (genes, neurons, voxels, or chemical constituents). Much of recent work has focused on understanding the computational complexity of proposed methods for Big Data. Sample complexity, however, has received relatively less attention, especially in the setting when the sample size n is fixed, and the dimension p grows without bound. To address this gap, we developed a unified statistical framework that explicitly quantifies the sample complexity of various inferential tasks. Sampling regimes can be divided into several categories: 1) the classical asymptotic regime where the variable dimension is fixed and the sample size goes to infinity; 2) the mixed asymptotic regime where both variable dimension and sample size go to infinity at comparable rates; and 3) the purely high-dimensional asymptotic regime where the variable dimension goes to infinity and the sample size is fixed. Each regime has its niche but only the latter regime applies to exa-scale data dimension. We illustrated this high-dimensional framework for the problem of correlation mining, where it is the matrix of pairwise and partial correlations among the variables that are of interest. Correlation mining arises in numerous applications and subsumes the regression context as a special case. We introduced a unified perspective of high-dimensional learning rates and sample complexity for different structured covariance models and different inference tasks. These correlation mining principles were extended to the case of complex valued random variables and, more specifically correlation mining in the spectral-domain, in a recent book chapter (H. Firouzi, D. Wei, and A.O. Hero, "Spectral correlation screening," in Excursions in Harmonic Analysis, Eds. R. Balan, M. Begue, J. J. Benedetto, W. Czaja and K. Okoudjou, Springer 2014).

2.2 Predictive correlation screening with resource constraints

We introduced a new approach to network variable selection, called Predictive Correlation Screening (PCS), for predictor design from a few samples. Predictive Correlation Screening implements false positive control on the selected variables, is well suited to small sample sizes, and is scalable to high dimensions. We established asymptotic bounds for Familywise Error Rate (FWER) and obtained bounds on resultant prediction mean square error. Unlike other variable selection methods based on prediction, e.g., PCA regression, lasso or marginal regression, PCS is highly scalable and has

good performance in the small sample regime. PCS can be motivated by the following two-stage predictor design problem for predictive health. In this problem we want to construct a simple predictor function that can accurately assess a subject's future health state using molecular (gene expression) data. The designer must learn this multivariate predictor based on assays of successive biological samples, which may be expensive to obtain and process. To save costs the designer adopts a two stage strategy. She assays the whole genome on a few samples and from these assays she selects a small number of variables using our theory of Predictive Correlation Screening. She subsequently performs a much cheaper set of assays using only the small number of selected variables on the remaining samples, to learn the predictor coefficients. Our PCS theory and experiments establish the superiority of Predictive Correlation Screening relative to LASSO and correlation learning in terms of MSE prediction performance and computational complexity. Our work on predictive correlation screening has been published in (H. Firouzi, B. Rajaratnam, A. Hero, "Predictive Correlation Screening: Application to Two-stage Predictor Design in High Dimension," AISTATS 2013) and in (H. Firouzi, B. Rajaratnam, A.O. Hero, "Two-stage variable selection for molecular prediction of disease," Proceedings of IEEE CAMSAP 2013). A journal version of this paper has been submitted (H. Firouzi, B. Rajaratnam, A.O. Hero, "Two-stage Sampling, Prediction and Adaptive Regression via Correlation Screening (SPARCS)," arxiv 1502:06189, Feb 2015).

2.3 Local hub screening in a correlation network

We have developed theory for controlling errors in localizing hubs, i.e., highly connected nodes, by performing hub screening of a correlation network. Hub screening is a method for discovering highly connected nodes in a large network. It is based on detecting nodes whose sample correlation, or partial correlation, with other nodes exceeds a user-defined threshold. Previous hub screening theory provided a way to select the correlation threshold in order to control the number of false positives globally across all nodes of the network. In the past year we have developed new local hub screening theory that can be used to control the false positives at a particular node. The significance of the local hub screening theory is that it allows one to control errors on *localization* of hubs where previous theory only controlled errors on *detection* of the presence of a hub somewhere in the network. The theory is related to previous global hub screening theory that established a Poisson-type limit to specify p-values on the number of spurious hub nodes found in the network. Local hub screening theory also establishes Poisson limits. However, instead of being on the global number of hub nodes found, here the Poisson limit applies to the node degree found at an individual node. This yields asymptotic p-values that are local to each node. We obtain convergence rates for proposed local hub screening method that are at least a factor of p faster than those of global correlation hub screening. The theory of local hub screening is reported in (Firouzi, Rajaratnam, and Hero, "Local hub screening for partial correlation graphs," Proceedings of SPIE workshop on Wavelets and Sparsity, 2013).

2.4 Non-convex sample-starved estimation of sparse inverse covariance matrices

The most popular methods for estimation of the inverse covariance (precision) matrix has been to impose sparsity and use ℓ_1-relaxation of the ℓ_0-norm in a convex penalized maximum likelihood framework. This has led to the widespread use of LASSO and GLASSO approaches to inverse covariance matrix estimation. These approaches rely on the accuracy of the convex ℓ_1-relaxation of the non-convex ℓ_0 penalized maximum likelihood problem. However, the use of ℓ_1 introduces

considerable shrinkage bias into the inverse covariance estimate. We have gone back to the basics, motivated by the fact that the most natural sparsity promoting norm is the nonconvex ℓ_0 penalty, and have proposed taking second look at the ℓ_0-penalized maximum likelihood problem despite its lack of convexity. As reported in our journal publication (G. Marjanovic and A. O. Hero, "l0 Sparse Inverse Covariance Estimation," IEEE Trans on Signal Processing, vol. 63, no. 12, pp. 3218-3231, May 2015), we have developed an tractible and scalable approach for solving the original nonconvex ℓ_0-penalized log-likelihood inverse covariance estimation problem without relaxation of the ℓ_0 penalty. We introduced a novel cyclic descent algorithm for this non-convex optimization problem with guaranteed convergence to a local minimizer. Due to the non-convexity of the objective function, the convergence analysis is highly nontrivial. It demonstrates the correctness of our proposed algorithm and the convergence properties were used to improve the cyclic descent algorithm. Simulations demonstrated the reduced bias and superior quality of our non-convex ℓ_0 penalized algorithm as compared to the standard convex GLASSO ℓ_1 penalized approach.

2.5 Positivity invariance of thresholded correlation matrices

We have established that soft thresholding of full rank correlation matrices preserves positive definiteness with high probability. Furthermore, soft thresholding of rank deficient correlation matrices, as occurs in the sample starved regime, restores positive definiteness with high probability. This is not true for hard thresholding. Estimation of covariance matrices in modern applications often requires some form of regularization. One of the most common approaches to high-dimensional covariance estimation is to induce sparsity via ℓ_1-regularization of inverse correlations, leading to classes of models popularly referred to as graphical models. Such methods provide a natural extension of the maximum likelihood estimation framework and are guaranteed to provide estimates which lie in the desired parameter space, namely the cone of positive definite matrices. Computing such sparse estimates, however, requires solving optimization problems. Although such approaches work well in moderate dimensions, they are not immediately scalable to the ultra high-dimensional settings necessitated by modern applications. A lesser-known approach is to obtain sparse estimates via thresholding of the individual elements of the sample covariance matrix by shrinking or setting them to zero, using a prescribed function (such as hard and soft thresholding). Such estimators have good asymptotic properties and are non-iterative, and therefore scale very well to high-dimensional settings. However, it remains unclear whether, in finite sample settings, thresholding methods produce positive definite covariance estimates. A non-positive definite estimator is of limited use for many downstream applications. Recent algebraic work by co-PI Rajaratnam builds on previous work by analysts Rudin and Schoenberg and shows that positive definiteness of such estimators can only be guaranteed in restrictive settings. In particular, the functions which leave the cone invariant are those which are analytic and absolutely monotonic. The work by co-PI Rajaratnam proceeds to relax the problem with modern motivations in mind. More specifically, in a series of papers in the Transaction of the American Mathematics Society, co-PI Rajaratnam together with his research group members demonstrates that restrictive assumptions can be removed when a) rank constraints are imposed (as necessitated by modern sample starved applications), and when b) sparsity constraints are imposed on the initial p.s.d matrix (as given by access to domain specfic knowledge). We have also empirically demonstrated that elementwise soft thresholding (compared to other thresholding procedures like hard thresholding) in fact retains positive definiteness in finite samples with extremely high probability, leading to viable estimators for prac-

tical applications. We show that it is possible to identify a priori a minimum level of regularization that will almost always yield a positive definite estimate. We then apply soft thresholding in several applications and observe that it is not only highly competitive, but also superior in terms of computational complexity. Soft thresholding can therefore be applied in high dimensional regimes where optimization-based methods are impractical. An article describing this work is being revised for resubmission (D. Guillot, A.O. Hero, B. Naul, and Rajaratnam B. Scalable Sparse Covariance Estimation via Soft Thresholding, 2015. (in revision).)

2.6 Applications to materials science

Our correlation mining principles and approaches were developed for several practical applications in materials science, in collaboration with Jeff Simons at the Air Force Research Laboratory, at Wright-Patterson AFB. These applications include a new physics-based dictionary approach to materials indexing using scanning electron microscopy (SEM), publised in (Y.-H. Chen, S.U. Park, D. Wei, M. Jackson, G. Newstadt, J. Simmons, M. De Graef and A. O. Hero, "A Dictionary Approach to the EBSD Indexing Problem," Microscopy and Microanalysis, vol. 21, no. 3, pp. 739-752, June 2015). This approach relies on computing correlations between Kikuchi patterns on a physical materials sample and the a dictionary of Kikuchi patterns computed from the transport physics associated with SEM system geometry and the crystal symmetry group associated with the orientations of polycrystaline grain structure in the sample. The bi-partite correlation graph obeys the principles developed our paper (A.O. Hero and B. Rajaratnam, "Foundational principles for large scale inference: Illustrations through correlation mining," IEEE Proceedings.vol. 105, no. 1, pp. 93-110, Jan. 2016). In SEM indexing of polychrysatline materials the Kikuchi patterns are indexed by the crystal orientation and the indices of the elements of the dictionary with highest correlation to a sample Kikuchi pattern is used to specify an estimate of the orientation. This is a suboptimal method of indexing the crystal orientations across the sample since it does not properly account for the symmetry group that governs unambiguous orientations. To address this we developed an optimal method of indexing that specifically incorporates the symmetry group in a maximum likelihood framework (Y.-H. Chen, D. Wei, G. Newstadt, M. Jackson, J. P. Simmons, M. De Graef and A. Hero, "Parameter estimation in spherical symmetry groups," IEEE Signal Processing Letters, vol. 22, no. 8, pp. 1152-1155, Jan. 2015).

3 Personnel Supported

Personnel supported

- Al Hero (UM faculty PI)
- Bala Rajaratnam (Stanford faculty co-PI)
- Joseph Romano (Stanford faculty collaborator)
- Apoorva Khare (Stanford Research Associate)
- Douglas Sparks (Stanford postdoc)
- Brett Naul (Stanford student)

- Taposh Banerjee (UM postdoc)
- Hamed Firouzi (UM student)
- Pin Yu Chen (UM student)
- Yu Hui Chen (UM student)
- Kristjan Greenewald (UM student)
- Brandon Oselio (UM student)

4 Publications

Refereed journals

1. Y.-H. Chen, S.U. Park, D. Wei, M. Jackson, G. Newstadt, J. Simmons, M. De Graef and A. O. Hero, "A Dictionary Approach to the EBSD Indexing Problem," Microscopy and Microanalysis, vol. 21, no. 3, pp. 739-752, June 2015.
2. Y.-H. Chen, D. Wei, G. Newstadt, M. Jackson, J. P. Simmons, M. De Graef and A. Hero, "Parameter estimation in spherical symmetry groups," IEEE Signal Processing Letters, vol. 22, no. 8, pp. 1152-1155, Jan. 2015.
3. D. Guillot and B. Rajaratnam, "Functions preserving positive definiteness for sparse matrices", Transaction of the American Mathematical Society - TAMS. (in print), 2013. Available as arxiv:1210.3894.
4. A.O. Hero and B. Rajaratnam, "Foundational principles for large scale inference: Illustrations through correlation mining," IEEE Proceedings.vol. 105, no. 1, pp. 93-110, Jan. 2016.
5. G. Marjanovic and A. O. Hero, "l0 Sparse Inverse Covariance Estimation," in IEEE Trans on Signal Processing, vol. 63, no. 12, pp. 3218-3231, May 2015.
6. G. Newstadt, E. Zelnio, and A.O. Hero III, "Moving target inference with Bayesian models in SAR imagery", IEEE Trans. on Aerospace and Electronic Systems, vol.50, no.3, pp.2004,2018, July 2014.
7. B. Rajaratnam, S. Roberts, D. Sparks, and O. Dalal. Lasso Regression: Estimation and Shrinkage via Limit of Gibbs Sampling. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2016. (to appear), http://arxiv.org/abs/1405.3034.
8. D. Guillot, B. Rajaratnam, and J. Emile-Geay. Statistical paleoclimate reconstruc- tions via Markov random fields. Annals of Applied Statistics, 9(1):324352, 2015. http://arxiv.org/abs/1309.6702.
9. K. Khare, SY. Oh, and B. Rajaratnam. A convex pseudo-likelihood framework for high dimensional partial correlation estimation with convergence guarantees. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 77(4):803825, 2015. http://arxiv.org/abs/1307.5381.
10. D. Guillot, A. Khare, and Rajaratnam B. Preserving positivity for rank-constrained matrices. Transactions of the American Mathematical Society (to appear), 2017. http://arxiv.org/abs/1406.0042.

11. B. Rajaratnam and D. Vincenzi. A note on covariance estimation in the unbiased estimator of risk framework. Journal of Statistical Planning and Inference, 2016. (to appear).

12. D. Guillot, A. Khare, and Rajaratnam B. Critical exponents of graphs. Journal of Combinatorial Theory: Series A , 139:3058, 2016. http://arxiv.org/abs/1504.04069.

13. D. Guillot and B. Rajaratnam. Functions preserving positive definiteness for sparse matrices. Transactions of the American Mathematical Society - TAMS, 367:627649, 2015. http://arxiv.org/abs/1210.3894.

14. D. Guillot, A. Khare, and Rajaratnam B. Complete characterization of Hadamard powers preserving Loewner positivity, monotonicity, and convexity. Journal of Mathematical Analysis and Applications, 425:1:489507, 2015. http://arxiv.org/abs/1311.1581.

15. P. Diao, D. Guillot, A. Khare, and Rajaratnam B. Differential calculus on graphon space. Journal of Combinatorial Theory: Series A, 133:183227, 2015. http://arxiv.org/abs/1403.3736.

16. D. Guillot, A. Khare, and Rajaratnam B. Preserving positivity for matrices with sparsity constraints. Transactions of the American Mathematical Society (to appear), 2016. http://arxiv.org/abs/1406.3408.

17. D. Guillot, A. Khare, and B Rajaratnam. The critical exponent conjecture for powers of doubly nonnegative matrices. Linear Algebra and its Applications, 439:24222427, 2013. http://arxiv.org/abs/1303.4701.

18. B. Rajaratnam and D. Vincenzi. A theoretical study of steins covariance estimator. Biometrika, 2015. (in revision).

19. O. Dalal and B Rajaratnam. G-AMA: Sparse Gaussian Graphical Model Estimation via Alternating Minimization. Biometrika, 2015. (in revision), http://arxiv.org/abs/1405.3034.

20. D. Horton, N.C. Johnson, D. Singh, D. Swain, B. Rajaratnam, and N. Diffenbaugh. Detection of robust trends in atmospheric circulation patterns and their contribution to extreme temperature occurrence. Nature, 522:465469, 2015.

21. B. Rajaratnam, J. Romano, M. Tsiang, and N. Diffenbaugh. Debunking the climate hiatus. Climatic Change, 133:2:129140, 2015.

22. M Tingley, P.F. Craigmile, M. Haran, B. Li, E. Mannshardt, and B. Rajaratnam. On discriminating between gcm forcing configurations using bayesian reconstructions of late-holocene temperatures. Journal of Climate, 28:82648281.

23. K. Rahman, S. Gorelick, J. Dennedy-Frank, Yoon. J., and B. Rajaratnam. Declining rainfall and regional variability changes in Jordan. Water Resources Research, 51:5:38283835, 2015.

Refereed conferences

1. Y.H. Chen, D. Wei, G. Newstadt, M. DeGraef, J. Simmons, A.O. Hero, "Statistical Estimation and Clustering of Group-invariant Orientation Parameters," Fusion 2015, Washington D.C 2015.

2. Y.H. Chen, D. Wei, G. Newstadt, J. Simmons, A.O. Hero, "Coercive Region-level Registration for Multi-modal Images," to appear at IEEE Intl Conf on Image Processing (ICIP), Quebec 2015.

3. P.-Y. Chen and A.O. Hero, "Multi-centrality graph PCA and its application to cyberintrusion detection," IEEE Intl Conference on Acoust, Speech and Signal Processing (ICASSP16), Shanghai 2016.

4. H. Firouzi, B. Rajaratnam, A.O. Hero, "Two-stage variable selection for molecular prediction of disease," Proceedings of IEEE CAMSAP 2013. (Nominated for Best Student Paper Award).

5. H. Firouzi, B. Rajaratnam, A.O. Hero, "Local hub screening for partial correlation graphs," Proceedings of SPIE workshop on Wavelets and Sparsity, 2013.

6. H. Firouzi, B. Rajaratnam, A. Hero, "Predictive Correlation Screening: Application to Two-stage Predictor Design in High Dimension," AISTATS 2013. Nominated for Best Student Paper Competition.

7. S.Y. Oh, O. Dalal, K. Khare, and B. Rajaratnam. Optimization methods for sparse pseudo-likelihood graphical model selection. In NIPS - Neural Information Processing Systems Foundation (conference proceedings). Montreal, CA, 2014. http://arxiv.org/abs/1409.3768.

8. Rajaratnam M., B., and K. A theoretical model for the term structure of corporate credit based on competitive advantage. In European Financial Management Association Annual Conference, 2015.

Book chapters

1. H. Firouzi, D. Wei, and A.O. Hero, "Spectral correlation screening," in Excursions in Harmonic Analysis, Eds. R. Balan, M. Begue, J. J. Benedetto, W. Czaja and K. Okoudjou, Springer 2014.

2. A.O. Hero and B. Rajaratnam, "Large scale correlation mining for biomolecular network discovery," in Big Data Over Networks, Eds. S. Cui, A. Hero, T. Luo, J. Moura, Cambridge University Press, 2015.

Technical reports

1. T. Banerjee and A.O. Hero, "Quickest Detection for Changes in Maximal kNN Coherence of Random Matrices," arxiv 1508.04720, Aug 2015.

2. Y.-H. Chen, D. Wei, G. Newstadt, M. DeGraef, J. Simmons, A.O. Hero, "Statistical Estimation and Clustering of Group-invariant Orientation Parameters," arxiv 1503.04127, Mar. 2015.

3. Yu-Hui Chen, Se Un Park, Dennis Wei, Gregory Newstadt, Michael Jackson, Jeff P. Simmons, Marc De Graef, Alfred O. Hero. "A Dictionary Approach to EBSD Indexing," arxiv 1502:07436, Feb 2015. (.html) . This paper appeared in the Journal of Microscopy and MicroAnalysis, Mar. 2015.

4. S.-Y. Oh, F. Dewey E. Ashley, P.S. Tsao, R. Olshen and B. Rajaratnam, "A Graphical-network Model Framework for Case-control Studies with Application to Cardiovascular Biology," Technical Report, 2013.

5. Yu-Hui Chen, Dennis Wei, Gregory Newstadt, Jeffrey Simmons, Alfred hero "Coercive Region-level Registration for Multi-modal Images," arxiv 1502:07436, Feb 2015.

6. H. Firouzi, B. Rajaratnam, A.O. Hero, "Two-stage Sampling, Prediction and Adaptive Regression via Correlation Screening (SPARCS)," arxiv 1502:06189, Feb 2015.

7. A. O. Hero and B. Rajaratnam, "Foundational principles for large scale inference: Illustrations through correlation mining," Stanford University Dept of Statistics Technical Report 2015-13.

8. A.O. Hero and B. Rajaratnam, "Large scale correlation mining for biomolecular network discovery," Stanford University Dept of Statistics Technical Report 2015-02

9. B. Rajaratnam and D. Sparks. MCMC-based inference in era of Big Data: A fundamental analysis of the convergence complexity of high-dimensional chains. Annals of Applied Probability (under review), 2016. http://arxiv.org/abs/1508.00947.

10. A. Khare and B. Rajaratnam. Probability inequalities and tail estimates for metric semigroups. Annals of Probability (under review), 2015. http://arxiv.org/abs/1506.02605.

11. D. Montague and B. Rajaratnam. Graphical Markov models for infinitely many variables. Transactions of the American Mathematical Society (under review), 2015. http://arxiv.org/abs/1501.07878.

5 Interactions/Transitions

5.1 Participation/presentations

The following are invited conference presentations, conferences and seminars in which co-PI's presented work related to this grant.

1. A. Hero was plenary speaker at the Future Directions in Compressive Sensing and Sensing-Processing Integration, workshop at Duke University (sponsored by the Office of the Secretary of Defense (OSD)) Jan 2016, entitled "The need for new theory and new models."

2. A. Hero gave plenary speaker at the IEEE Workshop on Signal Processing and Education, Sundance UT, July 2015, entitled "Large scale correlation mining."

3. A. Hero gave plenary speaker at the IEEE International Conf. on Image Processing (ICIP), Paris, 2014 on "Correlation mining in image and video processing."

4. A. Hero gave plenary lecture at IEEE CAMSAP Conference in Dec. 2013 entitled "Small Sample Correlation Mining in Massive Data Sets."

5. A. Hero gave plenary lecture at the Network Theory Symposium of IEEE GlobalSiP in Dec. 2013 entitled "Modeling of interaction networks: challenges and emerging solutions."

6. A. Hero gave plenary lecture at the 15 Year Anniversary of the Center for Imaging Science at Johns Hopkins in May 2013 entitled "Correlation mining in computational biology: pitfalls and opportunities."

7. A. Hero gave lecture at the "Large Scale Machine Learning for Big Data" workshop at the Institut Henri Poincarré in Paris, May 2013, entitled "Correlation mining."

8. A. Hero gave lecture at the Complex Networks Workshop, Eindhoven Jan 2013. "High dimensional dependency network analysis with limited data."

9. A. Hero gave lecture at the DARPA Workshop on Big Data and Large-Scale Analytics March 2013 entitled "Correlation mining in massive data."

10. A. Hero gave seminar and colloquium talks on correlation mining at the following venues
    (a) Distinguished lecture at Univ. of Rochester in Sept 2013. "Correlation mining in massive data."
    (b) Distinguished lecture at Wayne State Univ. Computer Science Dept in Feb 2013. "High Throughput Correlation Screening and Variable Selection for Massive Data."
    (c) Distinguished lecture at Texas A&M Univ. in Sept 2013. "Correlation mining in massive data."
    (d) UIUC ECE Colloquium Sept 2013, Correlation mining in massive data."

11. B. Rajaratnam gave an invited presentation at the DIMACS Workshop on Geological data fusion: Tackling the statistical challenges of interpreting past environmental change, Novel high dimensional statistical methodology for multiproxy paleoclimate reconstructions, Rutgers University (Jan, 2013)

12. B. Rajaratnam gave an invited presentation at the Carnegie-Mellon University, Department of Statistics: (April, 2013)

13. B. Rajaratnam gave an invited presentation at the University of West Indies, Port of Spain, Department of Mathematics and Statistics: (April, 2013)

14. B. Rajaratnam gave a series of three invited lectures at the Mathematics of Climate Change Conference, Guanajuato, Mexico (July, 2013)

15. B. Rajaratnam gave an invited presentation at the Inaugural meetings of the Canadian Statistical Sciences Institute (CANSSI) , Waterloo Ontario (Aug, 2013)

16. B. Rajaratnam gave an invited presentation at the first Mathematical Congress of America meetings, Special session on "Graph and Network Analysis in Geosciences" , Guanajuato, Mexico (Aug, 2013).

17. B. Rajaratnam gave an invited presentation at the University of Florida, Department of Statistics: (Feb, 2014)

18. B. Rajaratnam gave an invited presentation at the Pennsylvania Sate University, Department of Statistics: (Mar, 2014)

19. B. Rajaratnam gave one of the keynote addresses at "Symposium on spatial-temporal statistics: methods and applications", hosted by the Department of Statistics, University of California at Davis. (April, 2014)

20. B. Rajaratnam presented work on covariance estimation applied to climate change at the SAMSI closing workshop for the Program on Low-dimensional Structure in High-dimensional Systems (LDHD) (May, 2014).

21. B. Rajaratnam gave an invited presentation at the University of Lancaster, department of mathematics and statistics, June 2014.

22. B. Rajaratnam gave an invited presentation at the fourth annual workshop on Understanding Climate Change from Data, NCAR Boulder, CO, June 2014.

23. B. Rajaratnam gave the plenary talk at the The American Instituite of Mathematics, workshop on positivity, graphical models and multivari- ate dependencies, October 2014.

24. B. Rajaratnam gave an invited presentation at Duke University, Department of Statistical Science, November 2014.

25. B. Rajaratnam gave one of the two invited presentations at the University of Florence, Department of Statistics, Mini-workshop in graphical models, December 2014.

26. B. Rajaratnam gave an invited presentation at the European University Institute, Florence, Mini-workshop in statistics, December 2014.

27. B. Rajaratnam gave an invited presentation at the Joint Mathematics Meetings 2015, San Antonio, January 2015.

28. B. Rajaratnam gave an invited presentation at the North Carolina State university, Raleigh, January 2015.

29. B. Rajaratnam gave an invited presentation at the North Carolina State university, Raleigh, February 2015.

30. B. Rajaratnam gave an invited presentation at Duke University, February 2015.

31. B. Rajaratnam gave an invited presentation at the University of Michigan, Ann Arbor, April 2015.

32. B. Rajaratnam gave an invited presentation at the University of California, Davis, March 2015

33. B. Rajaratnam gave an invited presentation at the University of Colorado,Denver, April 2015

34. B. Rajaratnam gave an invited presentation at theUniversity of Illinois at Urbana-Champaign, May 2015

5.2 Consultative and advisory functions

1. A. Hero serves on the National Academy of Sciences Committee on Applied and Theoretical Statistics (CATS), 2012-. Advises the National Academies on the use of statistics in science, engineering and technology. CATS organizes workshops for government agencies (currently organizing workshops for NOAA, NSF, NIH).

2. A. Hero participated in the DARPA Workshop on Big Data and Large-Scale Analytics in 2013. The organizer of the workshop was DARPA DSO PM Tony Falcone. This workshop had the objective of addressing the important current and especially future problems in data analysis, as well as novel mathematical directions that might lead to their solution. The participants were asked to help shape the direction of mathematical research at DARPA

3. B. Rajaratnam performed as young investigator at DARPA on developing rigorous foundation for network analysis.

4. B. Rajaratnam (w/ H. Massam, D. Guillot, A. Khare) is principal organizer of the upcoming AIM workshop on "Positivity, graphical models, and modeling of complex multivariate dependencies," sponsored by American Institute of Mathematics (Oct 2014).

5. B.Rajaratnam serves on the advisory Committee on Probability and Statistics in the Physical Sciences of the Bernoulli Society.

6. B.Rajaratnam serves on the presidential search Committee of the Bernoulli Society.

5.3 Technology Assists, Transitions, and Transfers

1. A. Hero worked with Air Force Research Laboratory (AFRL) on applying correlation mining methods developed in this grant to automated target recognition (ATR) and Materials Science. He visited AFRL in April 2013 and Dec 2013. AFRL POCs: Edmund Zelnio and Jeff Simmons.

2. A. Hero worked with Army Research Laboratory (ARL) on multicriteria network discovery and human-in-the-loop sensing and processing. ARL POCs: Brian Sadler and Lance Kaplan. He visited ARL in Jan 2013, Mar 2014, June 2015, and Sept 2015.

3. A. Hero worked with MIT Lincoln Laboratory on network analytics for social media and human-in-the-loop sensing and processing. MIT POCs: Kevin Carter and Ted Tsiligkaridis.

4. A. Hero worked with Los Alamos National Laboratory on ...

5.4 New discoveries, inventions, or patent disclosures

New discoveries are reported in published papers. There were no inventions or patent disclosures.

6 Honors/Awards

Awards received during the period of this grant

1. A. Hero received the 2015 Society Award, IEEE Signal Processing Society, the highest award bestowed by the IEEE Signal Processing Society, presented at 2016 IEEE Intl. Conference on Acoustics, Speech and Signal Processing in Shanghai China.

2. A. Hero received the 2013 Technical Achievement Award, IEEE Signal Processing Society, presented at 2014 IEEE Intl. Conference on Acoustics, Speech and Signal Processing in Florence Italy.

3. IEEE CAMSAP 2013 Best Student Paper Competition Award (2nd place) or a paper co-authored with A. Hero's former student Zhaoshi Meng and his former post-docs Dennis Wei and Ami Wiesel entitled "Marginal Likelihoods for Distributed Estimation of Graphical Model Parameters," 2013 IEEE Computational Advances in Multi-Sensor Adaptive Processing workshop, St Martins.

4. IEEE ICIP 2013 Best Paper Award, for a paper co-authored with A. Hero's former student Paul Shearer and colleague Anna Gilbert entitled "Correcting Camera Shake by Incremental Sparse Edge Approximation," at the 2013 IEEE Intl. Conf. on Image Processing, Melbourne Australia.

5. AISTATS 2013 Notable Paper Award for paper by A. Hero's former student Zhaoshi Meng, and former post-docs Dennis Wei and Ami Wiesel entitled "Distributed Learning of Gaussian Graphical Models via Marginal Likelihoods," 16th Internation Conference on Artificial Intelligence and Statistics 2013, Scottsdale AZ.

Keynote talks, plenary talks, and distinguished lectures during the period of this grant

1. A. Hero was plenary speaker at the Future Directions in Compressive Sensing and Sensing-Processing Integration, workshop at Duke University (sponsored by the Office of the Secretary of Defense (OSD)) Jan 2016.

2. A. Hero was plenary speaker at the IEEE Workshop on Signal Processing and Education, Sundance UT, July 2015.

3. A. Hero was keynote speaker at the Conference on Scale Space and Variational Methods in Computer Vision, Bordeaux May 2015.

4. A. Hero was plenary speaker at the IEEE International Conf. on Image Processing (ICIP), Paris, 2014

5. A. Hero was keynote speaker at IEEE International Telecommunications Symposium (ITS), Sao Paulo 2014.

6. A. Hero was keynote speaker at Conference on Quantitative Non-destructive Evaluation (QNDE), Boise 2014.

7. A. Hero was plenary speaker at UC Riverside NSF IGERT Workshop and Retreat, Lake Arrowhead CA, Dec 2013. "Extraction of bio-molecular expression patterns from massive data: from hyperspectral imaging to personalized medicine ,"

8. A. Hero was plenary speaker at IEEE CAMSAP Workshop, St Martin, Dec 2013. “Small sample community detection in massive data sets,”
9. A. Hero was keynote speaker at New Sensing and Statistical Inference Methods Symposium, IEEE GlobalSIP Conference, Dec 2013. “Resource constrained adaptive sensing,”
10. A. Hero was keynote speaker at Network Theory Symposium, IEEE GlobalSIP Conference, Dec 2013. “Spatio-temporal graphical models for high dimensional network data.”
11. A. Hero gave a Distinguished lecture at Univ. of Rochester in Sept 2013. ”Correlation mining in massive data.”
12. A. Hero gave a Distinguished lecture at Wayne State Univ. Computer Science Dept in Feb 2013. ”High Throughput Correlation Screening and Variable Selection for Massive Data.”
13. A. Hero gave a Distinguished lecture at Texas A&M Univ. in Sept 2013. ”Correlation mining in massive data.”
14. A. Hero gave plenary lecture at the 15 Year Anniversary of the Center for Imaging Science at Johns Hopkins in May 2013
15. B. Rajaratnam was one of the key invited speakers at the Mathematics of Climate Change Conference, Guanajuato, Mexico (July, 2013)
16. B. Rajaratnam was one of the key invited speakers at the Inaugural meetings of the Canadian Statistical Sciences Institute (CANSSI), Waterloo Ontario (Aug, 2013)
17. B. Rajaratnam received the NSF Career Award by the NSF’s Division of Mathematical Sciences (2014).

Lifetime Honors/Awards

1. A. Hero received the 2015 Society Award, IEEE Signal Processing Society, the highest award bestowed by the IEEE Signal Processing Society, presented at 2016 IEEE Intl. Conference on Acoustics, Speech and Signal Processing in Shanghai China.
2. A. Hero received the 2013 Technical Achievement Award, IEEE Signal Processing Society, presented at 2014 IEEE Intl. Conference on Acoustics, Speech and Signal Processing in Florence Italy.
3. A. Hero received the 2011 Rackham Distinguished Faculty Achievement Award, University of Michigan
4. A. Hero received the IEEE Third Millenium Medal 2000.
5. A. Hero received the Meritorious Service Award, IEEE Signal Processing Society, 1998.
6. A. Hero was elevated to Fellow of IEEE, 1997.
7. B. Rajaratnam received the NSF Career Award, 2014.
8. B. Rajaratnam received the UPS Foundation Award, 2012.

9. B. Rajaratnam received the DARPA Young Faculty Award, 2011.
10. B. Rajaratnam received the NSA - American Mathematical Society (AMS) Young investigator award in the mathematical sciences, 2010.

1.

1. Report Type

Final Report

Primary Contact E-mail
Contact email if there is a problem with the report.

careymrz@umich.edu

Primary Contact Phone Number
Contact phone number if there is a problem with the report

734-647-1813

Organization / Institution name

University of Michigan

Grant/Contract Title
The full title of the funded effort.

Sample-starved large scale network analysis

Grant/Contract Number
AFOSR assigned control number. It must begin with "FA9550" or "F49620" or "FA2386".

FA9550-13-1-0043

Principal Investigator Name
The full name of the principal investigator on the grant or contract.

Dr. Alfred O. Hero III

Program Manager
The AFOSR Program Manager currently assigned to the award

Dr. James Lawton

Reporting Period Start Date

02/01/2013

Reporting Period End Date

01/31/2016

Abstract

In this research project we developed correlation mining methods to answer the following fundamental question about complex networks: What are the fundamental limits on the amount of information that can be inferred about a network from a small number n of indirect empirical observations? In these terms, the overall objective was to develop algorithms and establish performance limits for mining information from correlation networks. The focus was on the
sample starved regime arises when the number of variables (columns of the correlation matrix) is of the same order or larger than the number of observations available to estimate or detect patterns in the matrix. A new framework was developed to answer the above question based on spherical Gram matrices for inferring dependency structure of large networks from limited and/or incomplete sample observations of network behavior. The geometrical and statistical properties of these matrices was studied in the finite sample regime and in the asymptotic limit as numbers of samples and/or nodes become large. These properties led to quantfication of fundamental performance tradeoffs and
gave insights into phase transitions and convergence rates for inferring dependencies in network data. The theory was applied to practical complex network inference tasks including: online prediction, network variable selection and error controlled topology discovery.

Distribution Statement

This is block 12 on the SF298 form.

Distribution A - Approved for Public Release

Explanation for Distribution Statement

If this is not approved for public release, please provide a short explanation. E.g., contains proprietary information.

SF298 Form

Please attach your SF298 form. A blank SF298 can be found here. Please do not password protect or secure the PDF The maximum file size for an SF298 is 50MB.

SF.298.FA9550_13_1_0043.pdf

Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF . The maximum file size for the Report Document is 50MB.

afosrUM_Stanford11_finalreport_final.pdf

Upload a Report Document, if any. The maximum file size for the Report Document is 50MB.

Archival Publications (published) during reporting period:

Changes in research objectives (if any):

Change in AFOSR Program Manager, if any:

Effective 13 JAN 2015, AFOSR Program Officer changed from Dr. Robert Bonneau to Dr. James Lawton, AFOSR/RTA (703) 696-5999, james.lawton.1@us.af.mil

Extensions granted or milestones slipped, if any:

AFOSR LRIR Number

LRIR Title

Reporting Period

Laboratory Task Manager

Program Officer

Research Objectives

Technical Summary

Funding Summary by Cost Category (by FY, $K)

	Starting FY	FY+1	FY+2
Salary			
Equipment/Facilities			
Supplies			
Total			

Report Document

Report Document - Text Analysis

Report Document - Text Analysis

Appendix Documents

2. Thank You

E-mail user

Apr 28, 2016 14:15:56 Success: Email Sent to: careymrz@umich.edu